DECOMPOSITION OF VARIATION IN DIRECTIONAL DATA

C.M. Anderson and C.F.J. Wu

IIQP Research Report
RR-92-10

December 1992

DECOMPOSITION OF VARIATION IN DIRECTIONAL DATA

C.M. Anderson and C.F.J. Wu

Department of Statistics and Actuarial Science
University of Waterloo
December 1992

ABSTRACT

Much of the data motivating the research on directional statistics has been observational in nature. When directional data from a factorial experiment is considered, a number of interesting issues arise. In this paper, we consider what definitions are suitable for measuring main effects, and which of these are used by the existing methods for analysis of variance for directional data. We propose a new decomposition of variation which improves upon the existing approaches. Using bootstrap methods to approximate the distribution of the new test statistic makes the approach feasible and provides a dramatic increase in power. Simulation results reinforce the superiority of the new approach and data from a factorial experiment on brake rotors illustrate its use.

Key Words: Directional data, variance decomposition, interactions, factorial experiment, bootstrap.

1 Introduction

Much of the present study of directional data has been based on data sets that are observational in nature (see [3] and [9]). A relatively new aspect of study is to consider designed experiments as they relate to directional data. In some industrial settings, the production of circular or cylindrical objects gives rise to data with attributes that are best measured by directional variables. Variation in these attributes can often be reduced through factorial experiments. Controlling and adjusting the direction of the sample by changing factor levels is a potential area of interest.

Two procedures for analyzing directional data currently exist: the Watson and Williams approach (1956) and the method suggested by Harrison, Kanji and Gadsden (1986) and subsequently developed by Harrison and Kanji (1988). Little has been written about the nature of the effects being measured and the situations in which use of the different strategies would be appropriate. With traditional non-directional data, the notions of main effects and interaction are universally accepted and have been discussed and expanded upon extensively. A similar approach to these same notions in directional data is needed for a better understanding of the quantities being measured. When the existing methods are examined in the context of factorial experiments, it can be shown that the extension of the Watson-Williams approach is not mathematically sound because the interaction terms which are approximately chi-squared can be negative, and that the method proposed by Harrison et al. does not measure the usual effect difference.

In Section 2, we consider the existing tools for examining a full factorial design with a directional response. In Section 3, precise definitions for measures of effects are enumerated, with some possible applications. In Section 4, a new location-only measure is proposed to respond to some of the weaknesses of the Watson-Williams extension. Sections 5 and 6 compare the new method with the existing ones through simulation and a numerical example using experimental data on the balancing of brake rotors.

1.1 Notation

We begin with a brief review of notation and fundamental quantities for directional data. For a sample of directional data located on a circle, $(\theta_1, \cdots, \theta_n)$, the resultant length is defined to be $R = (c^2 + s^2)^{\frac{1}{2}}$, where $c = \sum_i \cos\theta_i$ and $s = \sum_i \sin\theta_i$. The resultant direction for the

sample is defined to be

$$\overline{x}_0 = \begin{cases} \arctan(s/c), & \text{if c} > 0\ , \\ \pi + \arctan(s/c), & \text{otherwise.} \end{cases} \tag{1}$$

Since it is often convenient to quantify the dispersion with a linear scale, we define the dispersion from the mean direction, μ, by

$$D = \frac{1}{n}\sum_i (1 - \cos(\theta_i - \mu))\,. \tag{2}$$

Note that $0 \leq D \leq 1$. If we are interested in the within-sample variance, we can use the circular sample variance $S_0 = 1 - \bar{R}$, where $\bar{R} = R/n$ is the mean resultant length. $S_0 = 0$ means no variation in the observations, and $S_0 = 1$ means that there is no mean direction.

If we consider the one-way situation, some additional notation is required. In this case, the resultant vector for each group is obtained using the same procedure as outlined above for the overall resultant vector. Hence for p groups we obtain the following summary statistics: $\theta_{1\cdot}, \cdots, \theta_{p\cdot}$ and $R_{1\cdot}, \cdots, R_{p\cdot}$ with the "$\cdot$" indicating use of all the observations within that group. For the two-way situation, the resultant vectors for all of the rows, columns and cells are obtained, and summarized by their angles and lengths. Hence, $\theta_{i\cdot}$ and $\theta_{\cdot j}$ are the resultant angles for the ith row and the jth column, respectively, and R_{ij} is the resultant length of the ijth cell.

Two common choices for distributions on the circle (Mardia, 1972) are the wrapped Normal and the von Mises, but the latter is generally preferred for computational convenience. The von Mises distribution, $VM(\mu, k)$, with mean direction μ and concentration parameter k has density of the form:

$$f(\theta) = \frac{1}{2\pi I_0(k)} exp\{kcos(\theta - \mu)\},$$

where $I_0(k)$ is the modified Bessel function with $k \geq 0$ and $theta \in (-\pi, \pi)$.

If a distribution with heavier tails is required, the wrapped Cauchy with spread parameter ρ is sometimes used. It has a density function of the form:

$$g(\theta) = \frac{1}{2\pi}\ \frac{1 - \rho^2}{1 + \rho^2 - 2\rho\cos\theta},$$

for $0 \leq \rho \leq 1$ and $theta \in (-\pi, \pi)$.

2 Existing Methods for Analysis of Variance

2.1 Watson and Williams

Watson and Williams (1956) noted that an analysis of variance property similar to the linear case exists for directional data in two or more dimensions. For a sample of size N the total dispersion, $ND = N - C$, where $C = \sum_i cos(\theta_i - \mu)$. Analogous to the linear case, this can be further divided into the deviation of the observed sample from the sample mean direction and the deviation of the sample mean from the true mean. If the data is a sample from a von Mises distribution, i.e. θ_i from $VM(\mu, k)$, then various quantities of interest are approximately chi-squared.

Using the development in Mardia (1972), for moderately large k (say $k \geq 2$), the individual dispersion of the observation from the true mean is approximately chi-squared, i.e.,

$$\varepsilon_i = 2k(1 - \cos(\theta_i - \mu)) \sim \chi^2_1.$$

Throughout the paper we use "$\sim$" to denote approximation. Hence, for the total variation, we obtain

$$2k(N - C) = \sum_i \varepsilon_i \sim \chi^2_N.$$

Since the true mean is generally not known, we use

$$2k(N - R) \sim \chi^2_{N-1}$$

if k is large. This can be obtained by using $2k(R - C) = 2kR(1 - \cos \overline{x}_0) \sim \chi^2_1$ and the identity $2k(N - C) = 2k(R - C) + 2k(N - R)$ in conjunction with a form of Cochran's Theorem.

Analogous results are also available for directional data in p dimensions:

$$2k(N - C) \sim \chi^2_{(p-1)N},$$

$$2k(N - R) \sim \chi^2_{(p-1)(N-1)}.$$

For q samples with N_i observations per group, we further decompose the variation into within- and between-group terms. By using Cochran's Theorem, we obtain the following results for the within- and between-group sum of squares:

$$SS_W = 2k(N - \sum_{i=1}^{N_i} R_i) \sim \chi^2_{N-q},$$

$$SS_B = 2k(\sum_{i=1}^{N_i} R_i - R) \sim \chi^2_{q-1},$$

and the two quantities are approximately independent.

Therefore, for large values of k, the following test statistic is suggested for testing the null hypothesis of no difference between group means:

$$\frac{(N-q)\ SS_B}{(q-1)\ SS_W} \sim F_{q-1,N-q}.$$

For the distributional results to hold, the following assumptions should be satisfied:

1. The samples are drawn from a Von Mises distribution.

2. The overall and individual concentration parameters must be sufficiently large for the χ^2 approximations to hold, say, $k \geq 2$.

3. The concentration parameters for each sample are assumed to have the same value.

The third assumption might be suspect in many designed experiments, where the difference in variation for different factor levels is exploited in order to reduce variation in the process. However it is the second assumption which may be most problematic in many applications, since it requires the data to be closely clustered. Stephens (1992, p. 76) noted that for circular data, namely on a circle, "the vectors are typically more widespread, indicating a lower value of k".

Since the emphasis of this paper is on data collected from a factorial experiment, we now consider the modifications needed to apply the method described by Watson and Williams to this context. An extension to accommodate a nested two-way analysis was proposed by Stephens (1982). Harrison et al. (1986) pointed out that this technique does not extend without distributional difficulty to the multi-way crossed layout. For a two-way design, we decompose the total variation $2k(N - R)$ into the sum of four terms by a straightforward extension, we obtain

$$Total\ Variation = R.V. + C.V. + I.V. + E.V.$$

whose components are defined as follows:

Row Variation: $R.V. = 2k(\sum_{i=1}^{p} R_{i..} - R_{...})$,

Column Variation: $C.V. = 2k(\sum_{j=1}^{q} R_{.j.} - R_{...})$,

Interaction Variation: $I.V. = 2k(\sum_{i=1}^{p}\sum_{j=1}^{q} R_{ij.} - \sum_{i=1}^{p} R_{i..} - \sum_{j=1}^{q} R_{.j.} + R_{...})$,

Residual Variation: $E.V. = 2k(N - \sum_{i=1}^{p}\sum_{j=1}^{q} R_{ij.})$,

where $R_{ij.}$, $R_{i..}$ and $R_{.j.}$ are the cell, row and column resultant lengths, respectively.

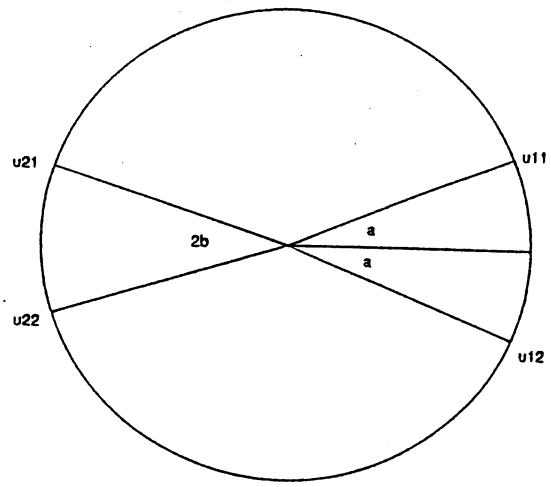

Figure 1: Example of Negative Interaction Term

For the one-way and the nested two-way classifications, the triangle inequality guarantees that both the between-group term and the residual term are non-negative. There is, however, no similar guarantee for the interaction term when we are considering the crossed layout. In fractional factorial designs, interactions are often aliased with main effects, so we must be concerned with the nature of the interaction terms. There is no mathematical reason to anticipate that the interaction terms will in fact be positive. We now consider a situation in which a negative interaction is obtained. Let $u11 = e^{i(a)}$, $u12 = e^{-i(a)}$ and $u21 = e^{i(\pi-b)}$, $u22 = e^{-i(\pi-b)}$. Then the following results are true:

$R_{jk} = 1$ for all j, k, $R_{1\cdot} = 2\cos a$, $R_{2\cdot} = 2\cos b$,

$R_{\cdot 1} = R_{\cdot 2} = 2\cos\left(\frac{\pi}{2} - \frac{a \pm b}{2}\right)$, and $R_{\cdot\cdot} = 2(\cos a - \cos b)$.

If we let $a = b$, then the interaction term becomes

$$4 - (4\cos a + 4\sin a) + 0 = 4(1 - \cos a - \sin a),$$

which is negative unless $a = 0$ or $\pi/2$, since $\cos a + \sin a \geq 1$. The case shown in Figure 1 corresponds to the case when $a \approx b$. Even in this simple situation with only four observations a negative interaction term can occur.

If the method is extended to deal with problems involving more factors, the interaction terms become increasingly complicated combinations of sums and differences of resultant lengths for different groupings.

In Table 2, we consider a 2^2 factorial model with ϕ_{ijk} denoting each angular measure of

Table 1: Combinations of Effects Used in Simulations

Run #	A_i	B_j	C_{ij}
1	0	0	0, 0.1, 0.25, 0.5, 1.0
2	0.25	0	0, 0.1, 0.25, 0.5, 1.0
3	0.5	0	0, 0.1, 0.25, 0.5, 1.0
4	1.0	0	0, 0.1, 0.25, 0.5, 1.0
5	0.25	0.5	0, 0.1, 0.25, 0.5, 1.0
6	0.5	0.5	0, 0.1, 0.25, 0.5, 1.0
7	1.0	0.5	0, 0.1, 0.25, 0.5, 1.0
8	0.25	1.0	0, 0.1, 0.25, 0.5, 1.0
9	1.0	1.0	0, 0.1, 0.25, 0.5, 1.0

the following form

$$\phi_{ijk} = \mu_0 + A_i + B_j + C_{ij} + \varepsilon_{ijk},$$

where the main effects, A_i and B_j, selected for these simulations vary between 0, ± 0.25, ± 0.5 and ± 1 radians, and the values of the interaction term, C_{ij}, are given in radians in the table. The error term, ε_{ijk}, is distributed as $VM(0, k)$ with k given in the table. It measures the angular deviation from the expected mean, and the values were generated using the algorithm in Best and Fisher (1979). Since in industrial settings it is generally expensive to obtain replicates, the methods were tested on small sample sizes. The results are based on nine sets of simulations (see Table 1), each of 1000 replications of the experiment at each level reported, with 20 data points, 5 at each combination of the factor levels. The first number reported in Table 2 is the average percentage of negative interactions terms for all the sets of simulations. The values in brackets give the range of values obtained. Repeating the simulations with a sample size of 40 did not yield substantially different results.

As can be seen, the number of negative interaction terms is considerable even when the concentration parameter is moderate. The negative interaction terms present a considerable problem when there is no interaction, or the interaction term is small. This situation is not avoided completely until both the size of the interaction and the concentration parameter are relatively large. The percentage of negative interactions rises when one main effect is large and the other relatively small (e.g. Simulation run 4 and 8). Hence the extension of the Watson-Williams approach to the factorial setting does not work well. Because of the

Table 2: Percentage of Analyses with Negative Interaction Terms

k	$C_{ij} = 0$	$C_{ij} = 0.1$	$C_{ij} = 0.25$	$C_{ij} = 0.5$	$C_{ij} = 1$
1	22.8 (12-37)	22.2 (13-37)	21.0 (8-37)	18.7 (6-39)	15.4 (3-34)
2	17.1 (5-30)	15.0 (5-26)	11.8 (3-26)	6.7 (0-20)	5.6 (0-23)
4	7.2 (0-15)	5.9 (0-13)	1.3 (0-3)	1.3 (0-7)	1.9 (0-4)
8	3.3 (0-8)	2.0 (0-5)	0.0	0.2 (0-2)	0.8 (0-7)
16	1.8 (0-5)	0.6 (0-2)	0.0	0.0	0.2 (0-2)

importance of interaction terms in designed experiment applications, this method proves unacceptable. However, it should be stressed that for one-way and nested two-way analyses, the Watson-Williams ANOVA decomposition does have appeal.

2.2 The L_2 Decomposition

Because of the difficulties involved in extending the Watson-Williams approach, Harrison et al. (1986) proposed a new method to measure distances between the mean resultant vectors of the cells, rows, columns and overall mean. The decomposition obtained is the usual L_2 distance for multidimensional data adapted to accommodate directional data. The sum of squares term measures the linear distance between the resultant vectors while incorporating the natural restriction that the observed data lie on the surface of the circle or sphere.

We take the two-way classification problem as a starting point for discussion. Consider a balanced experiment with p rows, q columns and m replicates in each combination of factors. Let u_{ijk} be the observation of unit length on the r-sphere corresponding to the kth replicate of the ith level of factor 1 and the jth level of factor 2, and $||u||^2 = u^T u$ to be the squared norm of u. For the 2-dimensional case, the data would correspond to observations in the x-y plane with $x_{ijk} = \cos\theta_{ijk}$, $y_{ijk} = \sin\theta_{ijk}$ and with the obvious restriction of $x_{ijk}^2 + y_{ijk}^2 = 1$.

The residual sum of squares can be re-expressed as a function of the resultant lengths

$$\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{m} ||u_{ijk} - \bar{u}_{ij\cdot}||^2 = N - \sum_i \sum_j \frac{R_{ij\cdot}^2}{m},$$

where $\bar{u}_{ij\cdot} = \sum_k u_{ijk}/m$ and $R_{ij\cdot}$ is the resultant length of the vectors with level i of factor 1 and level j of factor 2. If we define the mean resultant length of these same vectors to be

$\bar{R}_{ij.} = R_{ij.}/m$, the residual sum of squares takes the form $N - \sum_i \sum_j m\bar{R}^2_{ij.}$

A standard decomposition of the vector difference can be used to determine the main effects and the interactions, following Harrison et al. If we define $\bar{u}_{i..} = \sum_j \sum_k u_{ijk}/pm$, then we obtain the following decomposition:

$$\bar{u}_{ij.} - \bar{u}_{...} = (\bar{u}_{i..} - \bar{u}_{...}) + (\bar{u}_{.j.} - \bar{u}_{...}) + (\bar{u}_{ij.} - \bar{u}_{i..} - \bar{u}_{.j.} + \bar{u}_{...}).$$

The between-group sum of squares is decomposed using the L_2 distance to obtain sum of squares terms for the main and interaction effects:

$$\sum_i \sum_j m \left|\left|\bar{u}_{ij.} - \bar{u}_{...}\right|\right|^2 = \sum_i qm \left|\left|\bar{u}_{i..} - \bar{u}_{...}\right|\right|^2 + \sum_j pm \left|\left|\bar{u}_{.j.} - \bar{u}_{...}\right|\right|^2$$

$$+ \sum_i \sum_j m \left|\left|\bar{u}_{ij.} - \bar{u}_{i..} - \bar{u}_{.j.} + \bar{u}_{...}\right|\right|^2, \tag{3}$$

where $N_{ij} = m$, $N_{i.} = qm$, and $N_{.j} = pm$ are the number of observations in each cell, row and column respectively.

When these terms are simplified and re-expressed using the mean resultant lengths for the different groups, we obtain the following relation:

$$SS_T = SS_R + SS_C + SS_I + SS_{Res}, \tag{4}$$

where

$$SS_T = N - N\bar{R}^2_{..},$$

$$SS_R = \sum_i N_{i.}\bar{R}^2_{i.} - N\bar{R}^2_{..},$$

$$SS_C = \sum_j N_{.j}\bar{R}^2_{.j} - N\bar{R}^2_{..},$$

$$SS_I = \sum_i \sum_j N_{ij}\bar{R}^2_{ij} - \sum_i N_{i.}\bar{R}^2_{i.} - \sum_j N_{.j}\bar{R}^2_{.j} + N\bar{R}^2_{..},$$

and

$$SS_{Res} = N - \sum_i \sum_j N_{ij}\bar{R}^2_{ij},$$

are the total, row, column, interaction and residual sum of squares respectively.

Because each term is the sum of squared norms, all terms in the expressions (3) and hence (4) are non-negative. In addition, the approximate distribution for the main effect, interaction and residual sums of squares terms are all found to be approximately chi-squared

with the appropriate degrees of freedom for data from $VM(\mu, k)$ with sufficiently large k, say $k \geq 2$.

It should be noted that the expression (4) was originally developed by Harrison and Kanji (1988) from the geometrical perspective, but was expressed instead in terms of the net resultant lengths rather than the mean lengths. Our alternative development of (4) via the vector notation in (3) reveals its connection with the squared distance.

Hence, the statistics that can be obtained for testing the main and interaction effects are as follows:

Row Effect:

$$\frac{(SS_R)/(p-1)}{(SS_{Res})/(pq(m-1))} \sim F_{(p-1),pq(m-1)},$$

Column Effect:

$$\frac{(SS_C)/(q-1)}{(SS_{Res})/(pq(m-1))} \sim F_{(q-1),pq(m-1)},$$

Interaction Effect:

$$\frac{(SS_I)/(p-1)(q-1)}{(SS_{Res})/(pq(m-1))} \sim F_{(p-1)(q-1),pq(m-1)}.$$

To obtain these distributional results, Harrison and Kanji used the same assumptions made by the Watson-Williams method.

For a small concentration parameter (ie. $k \leq 2$), Harrison and Kanji showed that the residual sum of squares term is not well approximated by a chi-squared distribution, and hence suggested alternate test statistics using the average mean resultant length for the cells. For example, for the row effect the test statistic is

$$\frac{2}{1-\rho^2}(SS_R) \sim \chi^2_{2(p-1)},$$

where $\rho = I_1(\hat{k})/I_0(\hat{k})$. The columns and interaction effects, the test statistics are similar.

The result of their simulations showed that this approach provides a good fit to the mean and variance of the chi-squared distribution for a wide variety of concentration parameter values between 0 and 2. As the size of the sample and the concentration parameter increase, the approximations become increasingly accurate. It should be noted that in order to obtain an estimate of k, replication for each combination of factors is still required.

Therefore, depending on the estimate of the concentration parameter, different test statistics are available. The L_2 decomposition approach provides a method for testing the hypothesis of no difference between groups or classes for all values of the concentration parameter.

However, as will be seen later, it does not calculate the effect sum of squares using the usual measure of location effect.

3 Definition of Location and Dispersion Measures

One of the important contributions made by Taguchi (1986) is to stress the importance of separating the influence of factors on the mean from influence of factors on the variance of a response variable. When we approach design of experiment with this in mind, we are able to gain a better understanding of the underlying process and its effect on the response. In directional data, relatively little is known about the interpretation of the sums of squares quantities obtained by the Watson-Williams and Harrison-Kanji methods. To use these techniques effectively, it is important to understand what they are measuring and when we can expect to find a zero effect.

We have three possible definitions of a main effect: Location-only, dispersion-only and combined location-dispersion measures. In some applications the objective of study is to determine the optimal combination of factor levels for achieving some predetermined target direction. This would be analogous to a significant main effect in traditional data if a "large" difference in the mean of the data for different levels of that factor exists relative to the variability found naturally in the data. In other applications the main interest is to contain the observations to as small an arc of the circle as possible with little concern for the mean direction, that is, to examine the variability of the response. Finally, it is sometimes the case that both of these goals are of interest. These definitions of main effects will serve for a variety of study purposes that may be encountered.

The criteria used to define a zero main effect for each of these three definitions can be described in terms of characteristics of the group mean resultant vectors, with angle $\theta_{i\cdot}$ and length $R_{i\cdot}$.

A suitable criterion for a location-only effect to be zero is that the angular means are equivalent for the groups, i.e. $\theta_{1\cdot} = \cdots = \theta_{p\cdot}$. In general, an exact zero effect is hardly ever achieved because of measurement error, so to evaluate whether there is a significant location effect for a particular factor, we compare how far apart the resultant angles are relative to the variation within groups.

For the dispersion-only measure, a zero effect should reflect that the variation within the groups are the same. Hence since the resultant lengths are a popular choice for measuring

angular dispersion, a zero effect requires that the lengths of the mean resultant vectors are equal, i.e. $R_{1\cdot} = \cdots = R_{p\cdot}$. To determine a significant effect we would determine if the differences in mean lengths are too large to be attributed to the natural variation of the data.

Finally, if we are interested in a combined location-dispersion measure we require a definition incorporating both types of differences between groups in a single measure. In this situation, to obtain a zero effect for a factor, the criteria would be that both the resultant angles and their corresponding lengths are equivalent. This is clearly the most stringent definition for the main effect measures.

The formulation of these three definitions is designed to provide a framework for discussion and evaluation of analysis of variance or decomposition strategies. Little has been said in the literature about what the sums of squares terms in the existing methods are quantifying, perhaps because they were assumed to be analogous to the linear situation.

The Watson-Williams decomposition of the total variation can be shown to be a location-only measure, and a zero effect is obtained only when the resultant vectors for the groups have the same angular measure. However, as discussed previously there are problems with this model being extended to a multi-way layout.

For the method developed by Harrison et al., a zero effect is achieved only when both the angles and lengths of the resultant vectors are the same for all levels of the factor. Hence, the L_2 approach uses a combined location-dispersion measure definition to decompose the total variation. While initially it may seem advantageous to have a measure that combines both the mean and spread of the data, it also has some shortcomings. The process of combining both attributes into a single measure prevents easy access to the exact nature of the differences between groups. In this way, the analysis of variance table generated by the L_2 method contains information which is difficult to discern. In most situations one would be interested in knowing what attributes were different between groups: location, dispersion or both.

4 Decomposition based on the New Location Measure

Because it is often desirable to work with a measure that isolates the differences in the locations of the response, a new location-only measure with well-defined interaction sum of squares terms is required.

We first consider a simple one-way analysis of p groups, each with N_i observations. The basis of this approach is to remove the dispersion aspect of the data, and focus solely on the directional components of the different groups. In order to help facilitate this, we define some additional notation. For any vector v, we define

$$\underline{v} = \frac{v}{||v||}$$

to be the unit vector in the direction of v. Recall that the length of the mean resultant vector was related to the circular variance of the sample. By adjusting the vector so that it again lies on the unit circle (or sphere), we have removed the dependence of dispersion from our measure.

We can define the treatment sum of squares to be

$$\begin{aligned} SS_t &= \sum_i \sum_j ||\underline{u}_{i\cdot} - \underline{u}_{\cdot\cdot}||^2 \\ &= 2\sum_i N_i \left(1 - cos(\theta_{i\cdot} - \theta_{\cdot\cdot})\right), \end{aligned}$$

where $\theta_{i\cdot}$ is the angle of $u_{i\cdot}$, and $\theta_{\cdot\cdot}$ is the angle of $u_{\cdot\cdot}$.

To determine the relative size of this treatment sum of squares, a comparable residual sum of squares term is defined,

$$\begin{aligned} SS_r &= \sum_i \sum_j ||u_{ij} - \underline{u}_{i\cdot}||^2 \\ &= 2\sum_i \sum_j (1 - cos(\theta_{ij} - \theta_{i\cdot})), \end{aligned}$$

where θ_{ij} is the angle of the ijth observation. Only in the particular case when $N_i = 2$, SS_r is equivalent to that defined by the Watson-Williams decomposition.

From these two quantities we can construct an "F-style" test statistic of the form

$$F_{obs} = \frac{SS_t}{SS_r}$$

to determine whether the difference between the group means is large compared to the natural spread of the data within groups. Because it would be desirable to reduce the number of assumptions required to apply this approach, no distributional assumptions about the nature of the original data have been made. Hence the above quantity does not have any known distributional properties that may be used directly to determine the relative size of the

group effect. To alleviate the need for a distribution result for the test statistic, a bootstrap approach is taken to obtain an approximate significance level for the test statistic. A good overview of the bootstrap is provided by Efron and Tibshirani (1986). The simple bootstrap relies on resampling of the data under the assumption of no difference between groups. Under this assumption, all the data is treated as if it came from a single population with a common mean, and hence in each of the B iterations (where B is typically 100, 500 or 1000), data for each of the groups is drawn from the entire data set. The resampling is employed in order to build an approximate distribution for the test statistics, for comparing the observed data results with those from its approximate distribution. We would therefore reject the hypothesis of no difference between the groups if the number of bootstrapped test statistic values more extreme than the one obtained from the data is small. For example, if the cut-off value was taken to be 5%, then we would reject the null hypothesis if less that 5% of the bootstrapped test statistic values exceeded the observed value.

Because of the orthogonal nature of factorial designs, we are able to extend this approach to obtain a straightforward sum of squares for an interaction effect which possesses some desirable properties. A surrogate effect for the interaction term can be defined by creating a dummy main effect. Since the columns of the design matrix are orthogonal, the testing of different effects will be uncorrelated and hence the separate testing of the various effects will give a representative measure of the main and interaction effects. We now consider a factorial example with 2 factors, each with 2 levels and m observations per cell. If u_{ijk} is the kth observation with factor 1 at level i and factor 2 at level j, then we obtain the following sums of squares term:

$$\begin{aligned} SS_A &= \sum_i \sum_j \sum_k ||\underline{u}_{i..} - \underline{u}_{...}||^2 \\ &= 4m \sum_i (1 - \cos(\theta_{i..} - \theta_{...})), \\ SS_B &= 4m \sum_j (1 - \cos(\theta_{.j.} - \theta_{...})), \\ SS_{AB} &= 4m \sum_{l=1}^{2} (1 - \cos(\phi_l - \theta_{...})), \end{aligned} \tag{5}$$

where ϕ_1 is the angular mean of $\theta_{11.}$ and $\theta_{22.}$, and ϕ_2 is the angular mean of $\theta_{12.}$ and $\theta_{21.}$. The residual sum of squares is defined in an analogous way to the 1-way situation,

$$SS_R = \sum_i \sum_j \sum_k \left\| \underline{u}_{ijk} - \underline{u}_{ij\cdot} \right\|^2$$
$$= 2 \sum_i \sum_j \sum_k (1 - \cos(\theta_{ijk} - \theta_{ij\cdot})) .$$

The test statistics for testing the null hypothesis of no main effects and interactions would have an "F" form,

$$F_A = \frac{SS_A}{SS_R}, \quad F_B = \frac{SS_B}{SS_R}, \quad F_{AB} = \frac{SS_{AB}}{SS_R}$$

To obtain approximate significance levels for these test statistics, bootstrapping is again employed. Here for increased computer efficiency, all three test statistics should be calculated simultaneously for each bootstrapped sample.

For a general 2^n factorial design, the interaction terms are developed by grouping the observations into two mutually exclusive groups based on the levels of the factors considered. For example, for a three-way interaction (of factors A, B and C), we would treat all of the observations with an odd number of high levels as one group and the remaining observations as the second group. In other words, group 1 would consist of observations with factor combinations HHH, HLL, LHL, and LLH, while group 2 would include HHL, HLH, LHH and LLL, where HLH means that factor A is at the high level, factor B low and factor C high. Once the observations have been divided into two groups, the format used for SS_{AB} can be used.

In the following section a comparison of the three decomposition methods is provided giving an indication of the relative performance of these strategies under different conditions.

5 Comparison of Methods

The three methods outlined in Section 4 are quite different in their intended objectives and properties. Both the new measure and the Watson-Williams extension (henceforth noted by WW) are location-only measures, while the L_2 decomposition method uses the combined location-dispersion definition. A major difference between the two existing methods and the proposed method (Location-Only = LO) is the reliance on distributional assumptions. Both the WW as well as the L_2 methods depend on the assumption that the original data comes from a von Mises distribution with concentration parameter at least moderately large

(ie. $k \geq 2$). Stephens (1992) noted that the chi-squared approximations underlying the WW method are not as accurate for circular data as they are for data on a sphere. Stephens (1969) suggested a correction to improve its distributional properties. If, for moderate values of k, say $2 \leq k \leq 10$, the usual test statistics are multiplied by the factor $(1+3/8\hat{k})$, where $\hat{k}$ is the maximum likelihood estimate of k under von Mises, it is found to improve the distributional accuracy. Similarly Harrison and Kanji (1988) developed the multiplicative correction factor $(1-1/5\hat{k}-1/10\hat{k}^2)$ for this approach that is similar in nature to the one devised by Stephens for the WW approach.

On the other hand, the alternate approach does not make any distributional assumptions about the nature of the observed data, but rather relies on a bootstrap approach to simulate the approximate distribution of the test statistic for each collection of observations. The extra computational effort is worthwhile because of the gains in robustness.

To study the relative performance of the various methods, a number of simulations were performed. Initially, the simplest situation was considered, namely, one way classification with two groups. The results in Tables 3 and 4 are based on 100 replications of generating 10 observations in each group from the von Mises and Wrapped Cauchy distributions respectively, where the mean differs by the amount specified in the table. The same observations were tested by all three methods and the results are summarized. We use the abbreviations L2 for the L_2 decomposition and LO for the new location-only method with $B = 100$. The correction factors given previously are incorporated into the two existing methods with the true value of k being used for simplicity. The cutoff value that was considered was the 95% level.

When the data come from a von Mises distribution, we see that the three methods are quite different. The two existing methods have better size for the test and are close to the anticipated 95% level. The size of the test is conservative for the F-distribution approximation for the WW and L_2 methods. The proposed test tends to slightly overestimate the significance, but is close to the target value.

If we look at the ability of the methods to identify differences in mean, the two existing methods have weak power, even in cases with substantial angular differences and moderate concentration parameters. The new method cannot be formally compared with the other two methods since it does not have the same size, but the table clearly shows that it is powerful in identifying cases with a small difference between the group means.

Next we consider data that is not distributed as von Mises. In this case we selected the

Table 3: Von Mises Data - Group Differences Correctly Identified

	% Acceptance			% Rejection					
	Difference = 0			Difference = $\pi/4$			Difference = $\pi/2$		
k	WW	L2	LO	WW	L2	LO	WW	L2	LO
2	98	97	93	3	2	53	2	2	94
4	96	96	94	3	3	86	4	11	100
8	97	96	92	1	1	99	42	54	100
16	95	96	91	16	21	100	100	100	100

Table 4: Wrapped Cauchy Data - Group Differences Correctly Identified

	% Acceptance			% Rejection					
	Difference = 0			Difference = $\pi/4$			Difference = $\pi/2$		
Scale	WW	L2	LO	WW	L2	LO	WW	L2	LO
1	100	100	97	0	0	10	0	0	24
0.5	100	100	95	0	0	30	0	0	75
0.1	100	100	94	7	8	98	39	46	100

Table 5: Von Mises Data - Group Differences Correctly Identified

	% Acceptance Difference = 0			% Rejection Difference = $\pi/4$			% Rejection Difference = $\pi/2$		
k	LO	WW(B)	L2(B)	LO	WW(B)	L2(B)	LO	WW(B)	L2(B)
2	93	93	93	50	51	52	95	95	95
4	92	92	93	83	83	81	100	100	100
8	96	95	93	100	100	100	100	100	100
16	93	93	93	100	100	100	100	100	100

wrapped Cauchy distribution with heavier tails than the von Mises. Once again the new location-only measure has approximately the same size for the test, while the WW and L_2 methods seem to vastly underestimate the F-distribution cutoff value. The new method also seems to outperform the other two methods in identifying differences in group means. In the case where the scale parameter is small (i.e., the data is more concentrated), the WW and L_2 methods still have very little power, but the new method performs with remarkable accuracy. Hence, we can see that both the existing methods rely heavily on the distributional nature of the data, and falter when this assumption is not met.

5.1 Bootstrapping as an Alternative

From the results of the previous simulations we can draw a number of conclusions about the existing methods. First, even with the correction factors, the significance levels for the analyses using distributional results do not seem to be entirely accurate. Second, when the data does not come from a von Mises distribution, the tests become extremely conservative. Finally, the power of the tests is very limited unless the differences in means and the concentration parameter are large. Hence, the distributional assumptions required for the determination of the significance level seem to be restrictive, often unattainable in practical applications, and do not give desired results even in situations where the data is of the correct form. without using the usual F-distributional assumptions.

For a more complete understanding of the relative strengths and weaknesses of a nonparametric approach using the bootstrap approximation, we consider the test statistics of the existing methods without using the usual F-distributional assumptions. There are several

interesting aspects to Table 5, which summarizes the comparison of the three test statistics when bootstrapped. First of all, when the existing methods are bootstrapped instead of relying on distributional assumptions, they perform comparably to the new location-only method. All of the bootstrapped methods have size slightly smaller than the expected 5% cutoff, but when the power is compared to the distributional values, we see that bootstrapping provides a substantial improvement. If the angular distance between groups is moderate and the concentration parameter not too small, all three bootstrapped statistics have good ability in identifying between-group differences. A similar set of simulations for Wrapped Cauchy data was performed and the results were comparable. The relative performance of the three bootstrapped test statistics were very similar for this data.

Next the performance of the statistics was considered in situations where the concentration parameters differed for the two groups. Table 6 shows the result of simulations where the difference in means were 0, $\pi/4$ and $\pi/2$ radians. We considered a number of different values for k, and found that the relative performance of the three bootstrapped methods was impressive. Even when there were quite large discrepancies in the concentration parameters, they are all good at correctly identify between-group differences. To illustrate the different definitions that the location-only approaches and the L_2 approach use, if the difference in concentration parameter is large, i.e. $k_1 = 2$ and $k_2 = 16$, we see that the L_2 method identifies this as a difference between groups more frequently than the other methods. This is reflected by a size of only 82% when there is no difference in means, and by increased power for the case when the difference in means is $\pi/4$.

Hence, from these simulations we conclude that the bootstrap approach for any of the test statistics outperforms the distributional method in a number of important areas. It is more robust to differences in concentration parameters. As well it has dramatically better power in identifying group differences when they are small to moderate for a variety of distributions.

From these simulations, it appears that the bootstrapped WW results and the new location measure are comparable. They seem to have approximately the same size and power in the situations considered. The new location measure has the advantage, however, that it is well-defined and easily interpreted for multi-way situations. Recall that the WW approach can give a negative sum of squares term for interaction effects. The new method will provide an accurate test of effects for factorial experiments, since in these situations the design matrix has orthogonal columns. Finally, we make a few other notes about the

Table 6: Von Mises Data - Group Differences Correctly Identified with Different Concentration Parameter Values

Group 1 k_1	Group 2 k_2	% Acceptance Difference = 0				
		New	WW (Boot)	WW	L2 (Boot)	L2
2	2	93	93	100	93	100
	4	87	89	100	89	100
	8	95	95	100	92	100
	16	92	93	100	82	100
16	4	96	96	100	96	100
	8	93	93	100	93	100
	16	93	93	100	93	100
		% Rejection Difference = $\pi/4$				
2	2	50	51	0	52	0
	4	65	63	0	64	0
	8	74	72	0	75	0
	16	83	79	0	90	0
16	4	98	98	0	99	0
	8	100	100	3	100	3
	16	100	100	14	100	12
		% Rejection Difference = $\pi/2$				
2	2	95	95	0	95	0
	4	98	98	4	98	1
	8	100	100	4	100	3
	16	100	100	2	100	2
16	4	100	100	36	100	25
	8	100	100	85	100	67
	16	100	100	99	100	99

bootstrapped WW extension. It is computationally more demanding than the new method in that it requires a more complicated algorithm for obtaining the interaction test statistics. In addition, since the various sums are similar in magnitude, and the final statistic may be several orders of magnitude smaller than the individual terms being added and subtracted, there may be reduced numerical accuracy because of a possible cancellation effect.

The new approach and the L_2 method are comparable, when the concentration parameters for the groups are similar. However, when a difference in spread exists for the groups, the latter incorporates this fact in their determination of group differences. This is generally undesirable for most experimental design situations.

Hence, from practical considerations, the three methods are similar in many one-way situations if the concentration parameters for each group are similar. However, when we consider multi-way designs or designs with differences in the concentration parameter for various groups, both the Watson-Williams and L_2 approaches falter. Therefore, the new measure is superior for factorial designs when we wish to consider only differences in angular means.

6 Example using Brake Rotor Data

In this section we consider a real example involving the balancing of brake rotors. The data set consists of the location of a corrective weight that is added to the rim of the rotor. There are 24 unique locations (at 15 degree intervals) at which the weight can be added and based on the production of the rotor there is a natural orientation for the zero angle. The data are provided in angular form in Table 7. Three factors were thought to influence the location of the imbalance: thickness of sand mold (A), position of mold (B) and type of gang core box (C). A 2^3 factorial design was run in each of 8 blocks. Using Bartlett's test for the homogeneity of concentration parameters, as described by Stephens (1982), it can be shown that there is no significant difference between the k_i's for the different groups. Three summary tables for the analysis procedures are provided: Tables 8 and 9 give the Watson-Williams extension and the L_2 analysis with their respective correction factors. Finally Table 10 reports the significance levels (S.L.) and rankings of effects for the three analyses that were performed with the bootstrap.

An examination of the tables reveals a number of interesting aspects:

Table 7: Brake Rotor Data

	GROUP NUMBER							
	1	2	3	4	5	6	7	8
Factor A (Thickness)	L	H	L	H	L	H	L	H
Factor B (Position)	L	L	H	H	L	L	H	H
Factor C (Type)	L	L	L	L	H	H	H	H
CAST								
1	240	30	180	180	225	225	255	180
2	270	180	345	240	180	150	240	345
3	270	90	330	180	0	135	75	30
4	270	135	180	135	195	165	195	345
5	255	165	195	180	285	150	165	150
6	255	165	120	180	195	180	195	195
7	75	45	60	105	120	300	285	30
8	255	120	345	150	180	180	150	195

Table 8: Watson and Williams Analysis

Source	D.F.	Sum of Squares	P($F_{(d.f.,49)} > T$)	Rank
Block	7	11.802	.0003	1
A	1	4.490	.0007	2
B	1	0.175	.48	7
C	1	0.090	.61	8
AB	1	0.652	.17	5
AC	1	1.842	.02	3
BC	1	0.399	.28	6
ABC	1	0.931	.11	4
Error	49	19.836		
Total	63	40.217		

Table 9: L2 Decomposition Analysis

Source	D.F.	Sum of Squares	$P(F_{(d.f.,49)} > T)$	Rank
Block	7	12.677	.008	1
A	1	3.871	.01	2
B	1	0.240	.52	8
C	1	0.280	.49	7
AB	1	1.255	.15	5
AC	1	2.3205	.05	4
BC	1	1.102	.17	6
ABC	1	3.220	.02	3
Error	49	30.198		
Total	63	55.162		

Table 10: Bootstrapped Analyses

	LO		WW		L_2	
Source	S.L.	Rank	S.L.	Rank	S.L.	Rank
Block	.011	1	.007	1	.004	1
A	.016	2	.008	2	.006	2
B	.565	7	.540	7	.710	8
C	.723	8	.695	8	.724	7
AB	.129	5	.305	4	.182	5
AC	.107	4	.074	3	.046	4
BC	.421	6	.503	6	.218	6
ABC	.063	3	.354	5	.014	3

1. The significance levels for the different approaches can vary greatly for different factor combinations. Most notably are the blocking variable, Factor A and the three way interaction. In addition, the distributional and bootstrapped significance levels for a given technique also vary dramatically. This reinforces the simulation results which showed very different success rates for the two approaches.

2. The relative ranking of the effects varies with the approaches, although there is general consistency in the gross ranking of effects. All the approaches agree that Block and "A" are the most influential. Next the interactions "AC", "AB" and "ABC" follow in influence, with the remaining three factors being highly non-significant. The ranking of these three factors varies with the different analyses.

3. The tendency of the WW approach to give negative interactions, or to underestimate the real level of the interaction is demonstrated by the consistently lower sum of squares values for the two- and three-way interaction terms. Similarly, the significance levels for the bootstrapped WW interactions are larger than for the other methods, reflecting the underestimation of these terms.

4. The relative ranking of the new method appears to be a combination of the existing WW and L_2 methods (without bootstrap). It is similar to both, but there are differences in the ranking of the "AC" and "ABC" interactions with the WW method and for the main effects "B" and "C" for the L_2 method.

5. The WW and the bootstrapped WW methods use the same values of the test statistics but do not produce the same ranking of effects. This serves to illustrate that there are real differences for the test statistics distributions. For this data set, the assumed F-distribution does not appear to be close to the estimate obtained by bootstrapping. On the other hand, the relative ranking of effects for the two L_2 analyses is the same.

Overall, the new method performs well in its ability to identify the significant effects. This real data set serves to illustrate some of the theoretical results developed earlier.

7 Concluding Remarks

When directional data comes from designed experiments, a new array of questions and concerns are brought to light. Of primary interest are the definitions used to measure effects

for different objectives of a study. Three potential definitions are considered: location-only, dispersion-only and a combination of the two. The criterion for each is described and possible situations where they may be of interest are outlined. A new location-only measure was defined and tested. It has a well-defined interaction term, is easily and suitably applied to factorial situations, and has an intuitively pleasing decomposition of variation. In a simulation study for a simple situation, it is comparable to the two existing methods, and will function well in a wide variety of practical applications, because it is flexible enough to deal effectively with data from a number of distributions.

8 Acknowledgements

Research for this paper was funded in part by NSERC and MRCO. The authors are grateful for this financial support and also wish to thank R.J. MacKay and General Motors of Canada for assistance in obtaining the data set. Special thanks to Michael A. Stephens for his comments.

References

[1] BEST, D.J. and FISHER, N.I. (1979), "Efficient Simulation of the von Mises Distribution," *Appl. Statist.*, 28, 152-157.

[2] EFRON, B., and TIBSHIRANI, R. (1986), "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy," *Statist. Sci.*, 1, 54-77.

[3] HARRISON, D., KANJI, G.K. and GADSDEN, R.J. (1986), "Analysis of variance for circular data," *J. Appl. Statist.*, 13, 123-138.

[4] HARRISON, D. and KANJI, G.K. (1988), "The development of analysis of variance for circular data," *J. Appl. Statist.*, 15, 197-224.

[5] MARDIA, K.V. (1972), *Statistics of Directional Data,* London: Academic Press.

[6] STEPHENS, M.A. (1962), "Exact and approximate tests for directions I & II," *Biometrika*, 49, 463-477, 547-552.

[7] STEPHENS, M.A. (1969), "Tests for the von Mises distribution," *Biometrika*, 56, 149-160.

[8] STEPHENS, M.A. (1972), "Multisample Tests for the von Mises distribution," *J. Am. Statist. Assoc.*, 67, 456-461.

[9] STEPHENS, M.A. (1982), "Use of the von Mises distribution to analyse continuous proportions," *Biometrika*, 69, 197-203.

[10] STEPHENS, M.A. (1992), "On Watson's ANOVA for Directions," *The Art of Statistical Science*, New York: Wiley and Son.

[11] TAGUCHI, G. (1986), *Introduction to Quality Engineering: Designing Quality Into Products and Processes,* Tokyo, Japan: Asian Productivity Organization.

[12] WATSON, G.S. and WILLIAMS, E.J. (1956), "On the construction of significance tests on the circle and the sphere," *Biometrika*, 43, 344-352.